Volume 10 Number 3 1982 Nucleic Acids Research

DNA sequence of an immediate-early gene (IE mRNA-5) of herpes simplex virus type I

Roger J.Watson and George F.Vande Woude

Laboratory of Molecular Oncology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 13 October 1981; Revised and Accepted 1 December 1981

ABSTRACT

We describe a 2560 base pair herpes simplex virus type 1 (HSV-1) DNA sequence containing the entire immediate-early mRNA-5 (IEmRNA-5) gene. The 3' and 5' termini of IEmRNA-5 were mapped within this DNA sequence by single-strand specific endonuclease protection experiments. The IEmRNA-5 gene contains DNA sequences from both the unique (U_S) and reiterated (TR_S/IR_S) regions of the HSV-1 DNA short component and is interrupted by a single intron mapping in TR_S/IR_S. A search of the transcribed DNA sequence revealed no initiator codon within TR_S/IR_S. The first ATG was located 6 bases into U_S sequences and this reading frame was open for a further 87 codons. A second longer open reading frame (316 codons) was also observed in the 3' transcribed region. The oligonucleotide sequences adjacent to the IEmRNA-5 termini are discussed in relation to those of the HSV-1 thymidine kinase gene and other genes transcribed by RNA polymerase II.

INTRODUCTION

The accumulation of virus transcripts in the cytoplasm of cells infected with herpes simplex virus type 1 (HSV-1) is temporally controlled (1-3). In the absence of protein synthesis there is predominant synthesis of the five immediate-early mRNA's (designated IEmRNA-1 through -5), for which the genome map locations and directions of transcription have been described (4-6). Synthesis of early and late HSV-1 mRNA's is dependent upon the function of at least one immediate - early gene product (7). Synthesis of late mRNA's is dependent also upon replication of the virus DNA (1-3,8,9).

While it is generally accepted that the appearance of HSV-1 mRNA's in the cytoplasm is controlled, the mechanism by which this is regulated is uncertain. One obvious mechanism is that synthesis of HSV-1 mRNA's may be regulated at the level of transcription, under the control of virus transcriptional promoters. This would require that the promoter of an immediate-early gene is active in the absence of virus protein synthesis (i.e., in cycloheximide-treated cells) whereas that of an early gene is not. It has been demonstrated that one of the HSV-1 early genes, the thymidine kinase

0305-1048/82/1003-0979$2.00/0

(TK) gene, is not transcribed in cycloheximide-treated cells infected with wild-type virus (12). However, insertion of a DNA fragment containing the 5' region of an immediate-early gene (IEmRNA-3) adjacent to the TK structural gene sequences resulted in transcription of the TK gene at the immediate-early stage (13). These experiments underscore the importance of the 5' regions of these genes in controlling expression. The entire sequence of the TK gene has been resolved, and the 5'DNA sequences involved in gene regulation have been partially identified (10,35). As a step in understanding the differences in activity of immediate-early and early HSV-1 transcriptional promoters, we have determined the complete DNA sequence of an immediate-early gene, that encoding IEmRNA-5. Comparison of the DNA sequences of the IEmRNA-5 and TK genes revealed both common and unique features.

MATERIALS AND METHODS

Labeling DNA fragments: The primary source of HSV-1 DNA fragments used in this study was plasmid pBR322 subclones (isolates pKL43 and pKL63) of a defective HSV-1 DNA Eco R1 fragment cloned in bacteriophage λgtWES (isolate 12-7; reference 14). Further DNA fragments were obtained from the HSV-1 Eco R1 H DNA fragment cloned in λgtWES (Dec 36; reference 15). Restriction endonuclease cleavage maps for the HSV-1 DNA sequences of the regions of overlap between 12-7 and Dec 36 were identical for every enzyme tested. Fragments prepared by restriction endonuclease digestion were 5' and 3' ^{32}P-labeled as described previously (16).

Nuclease analysis of RNA:DNA hybrids: The procedures for formation of hybrids between terminally ^{32}P-labeled DNA fragments and cytoplasmic immediate-early RNA and for nuclease S1 digestion have been described (17). Treatment of hybrids with mung bean nuclease (450 units per reaction:obtained from PL Biochemicals) was as described for S1 except that the time of incubation was 1 hr and the temperature was 37°.

DNA sequencing: Sequencing of end-labeled DNA fragments was performed essentially as described by Maxam & Gilbert (18). In place of the pyridinium formate reaction, however, was substituted a purine-specific reaction using 50 µl 88% formic acid (Baker Chemical Co.) added to 10 µl H_2O and 10 µl end-labeled DNA. This reaction (10-15 minutes at 20°C) was stopped as described for the hydrazine reactions (18). Products of the sequencing reactions were generally resolved on 8.3M urea - containing 8% or 20% polyacrylamide gels (18). On occasion, these products were resolved on 20% polyacrylamide gels containing 90% formamide (19).

RESULTS AND DISCUSSION

Nucleotide sequence of the IEmRNA-5 gene: The map location of the IEmRNA-5 gene was previously determined by nuclease analysis and electron microscopy (17). It was found that the 5' terminus of IEmRNA-5 maps in the BamH1 4/EcoR1 (B4R1) DNA fragment and is located approximately 500 base pairs (bp) from the EcoR1 cleavage site (Figure 1). The IEmRNA-5 gene contains a single intervening sequence mapping in B4R1 approximately 250 bp from the IEmRNA 5' terminus. The sequence and location of this 149 bp intron has been previously reported (16). Both the IEmRNA-5 5' terminus and the intron map within the inverted repeats (TR_S/IR_S) which flank the short unique (U_S) region of the virus genome. The IEmRNA-5 3' cotranscript extends across the 180 bp BamH1 5/BamH1 4 (B5B4) DNA fragment and terminates approximately 570 bases (b) from the BamH1 5 cleavage site within the Sal1/BamH1 5 (SalB5) DNA fragment (Figure 1). To identify any possible control elements 5' or 3' to the structural gene a continuous sequence extending from the EcoR1 site for 2560 bp was resolved. This sequence included the entire B4R1 and B5B4 fragments and approximately 620 bp of the SalB5 fragment (Figure 2).

The strategy used for sequence determination is illustrated in Figure 1 and involved construction of detailed restriction maps for Sma 1 and a number of other enzymes. Isolated DNA fragments were 3' or 5' ^{32}P-labeled, then cleaved with a second enzyme. The uniquely labeled DNA fragments were isolated by polyacrylamide gel electrophoresis, subjected to partial chemical degradation and the products resolved on sequencing gels. The sequences of both strands of the DNA were determined, as were the sequences across restriction sites utilized for labeling or secondary cleavage. In general, no discrepancies were noted between the sequences of the complementary strands. The occasional mismatch of complementary sequences was ascribed to two causes. First, methylation of BstN1 cleavage sites was observed. This problem was resolved by construction of a BstN1 cleavage map. Second, band compression was occasionally noticed on autoradiographs of the sequencing gels. This problem was resolved either by running the urea-containing polyacrylamide gels at a higher wattage and temperature (70°-75°) or by running the products of the sequencing reactions on formamide-containing polyacrylamide gels (19). Band compression in sequencing runs of the Sma1 D subfragment of B4R1 (Figure 1) was particularly difficult to resolve. The cause of this compression could be ascribed to two G:C -containing inverted repeats of 12 and 7 bp, each of which contain two A:T pairs at their centre of symmetry. These inverted repeats are located within residues 334-359 and 391-406 in the DNA

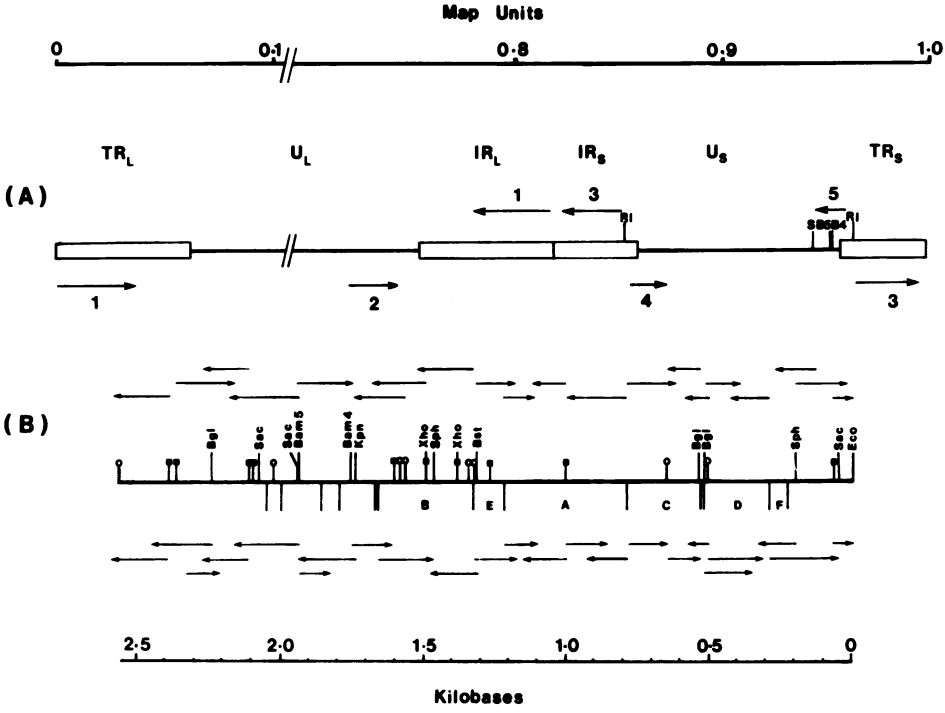

Figure 1. (A) Locations and directions of synthesis of the IEmRNAs (1 through 5) on the HSV-1 DNA. The locations of the SalI (S), BamHI (B4, B5) and EcoRI (RI) cleavage sites that define the SalB5, B5B4 and B4RI DNA fragments are indicated. (B) DNA sequencing strategy. Represented are restriction endonuclease cleavage maps of the DNA sequence shown in Figure 2 for the enzymes HinfI ( ), TaqI ( ), SmaI (|), BglI (Bgl), SacII (Sac), BamHI (Bam5, Bam4), KpnI (Kpn), XhoI (Xho), SphI (Sph), BstEII (Bst) and EcoRI (Eco). The letters below the line (A through E) refer to the designation of the SmaI subfragments of B4RI. Horizontal arrows indicate the limits and strand of the DNA sequenced in each of the various sequence determinations.

sequence (Figure 2).

Localization of the IEmRNA-5 5' terminus: The results of previous analyses (16,17) indicated that the IEmRNA-5 5' terminus maps within the 257 bp Sma1 C subfragment of B4R1 (Figure 1). This was confirmed by hybridizing the Sma1 C fragment, ^{32}P-labeled at each 5' terminus, to cytoplasmic immediate-early RNA. Nuclease S1 digestion of the resultant RNA/DNA duplexes led to the protection of a 200 b DNA fragment (data not shown), indicating that the mRNA 5' terminus maps approximately 200 bp from the Sma1 cleavage site near residue 590 (Figure 2).

To locate the 5' terminus more precisely a further nuclease analysis was

performed using the 117 bp Hinf1/Sma1 subfragment of the Sma1 C fragment (residues 534-650) 5'-^{32}P-labeled at the Hinf1 terminus. A fraction of the sample was hybridized to total cytoplasmic immediate-early RNA and the resultant hybrids were treated with either nuclease S1 or another single-strand specific endonuclease, mung bean nuclease. A sample of each digest was then electrophoresed on a sequencing gel in parallel with a sequencing run of the Hinf1/Sma1 fragment. As described by others (10, 19-21), this procedure enables the map location of the mRNA 5' terminus (the "cap" site) to be determined by correlation of the size of the fragments protected against nuclease digestion with the parallel sequence. Such correlation requires a -1b size correction for the nuclease digestion products to account for differences in the 3' termini remaining following chemical degradation as opposed to enzyme digestion (19-21). The results obtained are shown in Figure 3. Both with nuclease S1 and mung bean nuclease, protection of two predominant DNA fragments differing in size by 2b was observed. The presence of a ladder of additional minor bands obtained by nuclease digestion is generally observed with S1 and seems to be a characteristic of the enzyme rather than reflecting further mRNA terminal heterogeneity (20).

By reference to the parallel sequencing run, the probable complements of the capped IEmRNA-5 nucleotides were resolved to be T and C, corresponding to the complementary nucleotides A and G at positions 591 and 593 (Figure 2). While this method does not definitively identify the 5' terminal nucleotides of the transcript, for which direct sequence analysis of the mRNA is required, it does give results that are comparable to those obtained by the method involving sequence determination of reverse transcriptase extension products (11). It seems reasonable to conclude, therefore, that the IEmRNA-5 transcript is initiated in the region of residues 591-593. That the putative 5' nucleotides are purines mapping in the sequence CGCAC is consistent with the observation that initiation generally occurs in a sequence Pyrimidine-Purine-Pyrimidine (20,22).

The DNA sequences identified by the above nuclease analysis map in the short inverted repeat (TR_S/IR_S) regions of the HSV-1 genome. We have previously shown that IEmRNA-5 shares 5' sequences with IEmRNA-4, a consequence of their containing common TR_S/IR_S sequences (17). As such, the probes ostensibly used to map the 5' terminus of IEmRNA-5 also map that of IEmRNA-4. Since an unambiguous result of this analysis was obtained, the 5' termini of these two mRNA's are assumed to be identical. Sequence determination of the 5' region of the IEmRNA-4 gene (Watson, R. J., unpublished results)

GAATTCCATTATGCACGACCCCGCCCCGACGCCGGCACGCCGGGGGCCCGTGGCCGCGGC 60
EcoR1

CCGTTGGTCGAACCCCCGGCCCCGCCCATCCGCGCCATCTGCCATGGGCGGGGCGCGAGG 120

GCGGGTGGGTCCGCGCCCCGCCCCGCATGGCATCTCATTACCGCCCGATCCGGCGGTTTC 180

CGCTTCCGTTCCGCATGCTAACGAGGAACGGGCAGGGGGCGGGGCCCGGGCCCCGACTTC 240

CCGGTTCGGCGGTAATGAGATACGAGCCCCGCGCGCCCGTTGGCCGTCCCCGGGCCCCCG 300

GTCCCGCCCGCCGGACGCCGGGACCAACGGGACGGCGGGCGGCCCAAGGGCCGCCCGCCT 360

TGCCGCCCCCCCATTGGCCGGCGGGCGGGACCGCCCCAAGGGGGCGGGGCCGCCGGGTAA 420

AAGAAGTGAGAACGCGAAGCGTTCGCACTTCGTCCCAATATATATATATTATTAGGGCGA 480

AGTGCGAGCACTGGCGCCGTGCCCGACTCCGCGCCGGCCCCGGGGGCGGGCCCGGGCGGC 540

GGGGGGCGGGTCTCTCCGGCGCACATAAAGGCCCGGCGCGACCGACGCCCGCACACGGCG 600
• •mRNA5'

CCGGCCACGAACGACGGGAGCGGCTGCGGAGCACGCGGACCGGGAGCGGGAGTCGCAGAG 660
termini

GGCCGTCGGAGCGGACGGCGTCGGCATCGCGACGCCCGGCTCGGGATCGGGATCGCATCG 720

GAAAGGGACACGCGGACGCGGGGGGGAAAGACCCGCCCACCCCACCCACGAAACACAGGG 780

GACGCACCCCGGGGGCCTCCGACGACAGAAACCCACCGGTCCGCCTTTTTTGCACGGGTA 840

AGCACCTTGGGTGGGCGGAGGAGGGGGGGACGCGGGGGCGGAGGAGGGGGGGACGCGGGG 900

GCGGAGGAGGGGGGGACGCGGGGGCGGAGGAGGGGGGGACGCGGGGGCGGAGGAGGGGGC 960

TCACCCGCGTTCGTGCCTTCCCGCAGGAGGAACGCCCTCGTCGAGGCGACCGGCGGCGAC 1020

CGTTGCGTGGACCGCTTCCTGCTCGTCGGGAAAAGCATGTCGTGGGCCCTGGAAATGGCG 1080
TR_S | IR_S : U_S Met Ser Trp Ala Leu Glu Met Ala

GACACCTTCCTGGACAACATGCGGGTTGGGCCCAGGACGTACGCCGACGTACGCGATGAG 1140
Asp Thr Phe Leu Asp Asn Met Arg Val Gly Pro Arg Thr Tyr Ala Asp Val Arg Asp Glu

ATCAATAAAAGGGGGCGTGAGGACCGGGAGGCGGCCAGAACCGCCGTGCACGACCCGGAG 1200
Ile Asn Lys Arg Gly Arg Glu Asp Arg Glu Ala Ala Arg Thr Ala Val His Asp Pro Glu

CGTCCCCTGCTGCGCTCTCCCGGGCTGCTGCCCGAAATCGCCCCCAACGCATCCTTGGGT 1260
Arg Pro Leu Leu Arg Ser Pro Gly Leu Leu Pro Glu Ile Ala Pro Asn Ala Ser Leu Gly

GTGGCACATCGAAGAACCGGCGGGACCGTGACCGACAGTCCCCGTAATCCGGTAACCCGT 1320
Val Ala His Arg Arg Thr Gly Gly Thr Val Thr Asp Ser Pro Arg Asn Pro Val Thr Arg

TGAGTCCCGGGTACGACCATCACCCGAGTCTCTGGGCGGAGGGTGGTTCCCCCCCGTGGC 1380
***

TCTCGAGATGAGCCAGACCCAACCCCCGGCCCCAGTTGGGCCGGGCGACCCAGATGTTTA 1440

CTTAAAAGGCGTGCCGTCCGCCGGCATGCACCCCAGAGGTGTTCACGCACCTCGAGGACA 1500

CCCGCGCATGATCTCCGGACCCCCGCAACGGGGTGATAATGATCAAGCGGCGGGGCAATG 1560

TGGAGATTCGGGTCTACTACGAGTCGGTGCGGACACTACGATCTCGAAGCCATCTGAAGC 1620

CGTCCGACCGCCAACAATCCCCAGGACACCGCGTGTTCCCCGGGAGCCCCGGGTTCCGCG 1680

ACCACCCCGAGAACCTAGGGAACCCAGAGTACCGCGAGATCCCAGAGACCCCAGGGTACC 1740

GCGTGACCCCAGGGATCCACGACAACCCCGGTCTCCCAGGGAGCCCCGGTCTCCCCGGGA 1800
BamHl 4

```
GCCCCGGACCCCACGCACCCCCCGCGAACCACGTACGGCTCGCGGGTCTGTATAGCCCGG   1860
GCAAGTATGCCCCCCCTGGCGAGCCCAGACCCCTTCTCCCCACAAGATGGAGCATACGCTC  1920
GGGCCCGCGTCGGGATCCACACCGCGGTTCGCGTCCCGCCCACCGGAAGCTCAACCCACA  1980
          BamHl  5
CGCACTTGCGGCAAGACCCGGGCGATGAGCCAACCTCGGATGACTCAGGGCTCTACCCTC  2040
TGGACGCCCGGGCGCTTGCGCACCTGGTGATGTTGCCCGCGGACCACCGGGCCTTCTTTC  2100
GAACCGTGGTCGAGGTGTCTCGCATGTGCGCTGCAAACGTGCGCGATCCCCCGCCCCCGG  2160
CTACAGGGGCCATGTTGGGCCGCCACGCGCGGCTGGTCCACACCCAGTGGCTCCGGGCCA  2220
ACCAAGAGACGTCGCCCCTGTGGCCCTGGCGGACGGCGGCCATTAACTTTATCACCACCA  2280
TGGCCCCCCGCGTCCAAACCCACCGACACATGCACGACCTGTTGATGGCCTGTGCTTTCT  2340
GGTGCTGTCTGACACACGCATCGACGTGTTCGTACGCGGGGCTGTACTCGACCCACTGCC  2400
TGCATCTGTTTGGTGCGTTTGGGTGTGGGGACCCGGCCCTAACCCCACCCCTGTGCTAGG  2460
GCAATTTGTACCCTTAATAAATTTCACAAACAGATTTTATCGCATCGTGTCTTATTGCGG  2520
GGGAGAAAACCGATGTCGGCATAGAAAACCGCCATGATTC                       2560
```

Figure 2. DNA sequence of the non-coding strand of the IEmRNA-5 gene: the DNA is transcribed from left to right. Horizontal arrows indicate the inverted repeats in the SmaI D subfragment of B4RI. Also underlined are the putative "TATA" box and polyadenylation signal and the recognition sites for EcoRI and BamHI. The intron is shown bounded by parentheses and the junction between TR_S/IR_S and U_S is delineated by a dotted line. The sequence of the predicted protein product of this mRNA is indicated. Additional restriction endonuclease cleavage sites are present at the following positions: AccI (1572); AluI (1968); AosI (2057); AsuII (2098); AvrII (1694); BclI (1540); BstEII (1311); BstNI (1068, 1089, 1112, 1641, 1731, 1749, 1776, 1874, 2063, 2245); DdeI (2024); HaeI (2326); HaeII (493, 597, 2051); HinfI (505, 650, 1322, 1346, 1565, 1581, 2022, 2556); KpnI (1735); MboII (1271); RsaI (1119, 1129, 1331, 1709, 1736, 1833, 2372, 2384, 2468); Sau3A (167, 705, 711, 1140, 1510, 1541, 1600, 1718, 1754, 1934, 2145); SmaI (225, 289, 519, 531, 788, 1219, 1326, 1659, 1668, 1794, 1856, 1997, 2047); SacII (54, 1942, 2077); SphI (193, 1464); TaqI (68, 1001, 1269, 1383, 1492, 1604, 2099, 2110, 2361, 2388); XhoI (1382, 1491).

revealed identical sequence with that of the IEmRNA-5 gene. These latter analyses have allowed the junctions between TR_S/IR_S and U_S sequences to be resolved and that within the IEmRNA-5 gene (residues 1050-1051) is shown in Figure 2.

The DNA sequence 5' to the transcribed region of the IEmRNA-5 gene was compared to that of the HSV-1 TK gene (10,11) and of those described for other genes trancribed by RNA polymerase II (20-26). 20-30 b upstream of the IEmRNA-5 "cap site" is the sequence CACATAAA (residues 561-569) which is similar to the sequence CATATTAA found in the equivalent position of the TK gene. Presumably these sequences are related to the TATA box that is found in this position in many, though not all, RNA polymerase II transcribed genes (26) and which appears to be required for positioning the start of the transcribed

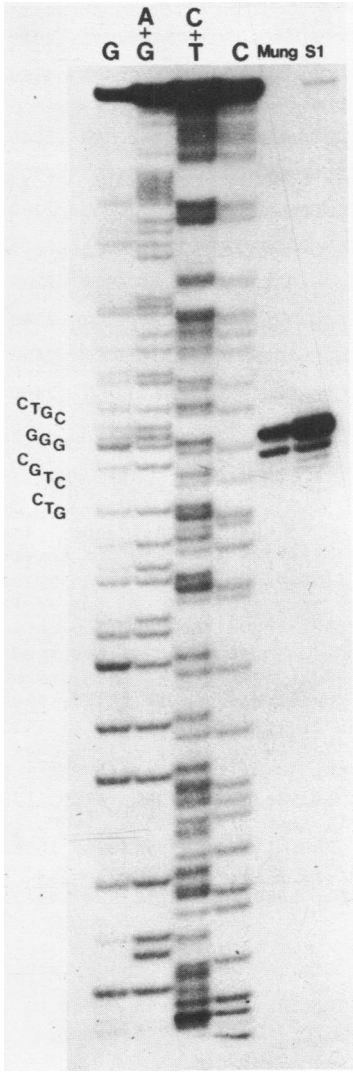

Figure 3. Identification of the 5' termini of IEmRNA-5. The products of mung bean and S1 nuclease digestion of hybrids formed between cytoplasmic immediate-early RNA and the 5' terminally labeled Hinf1/SmaI subfragment of SmaI C were electrophoresed in parallel to the products of the sequencing reactions of that DNA fragment.

region (23). Preceding the presumed IEmRNA-5 TATA box by 8b is the sequence GGTCT (residues 549-553) which is similar to the sequence GGTCC found in the equivalent position of the TK gene. It has been reported (25) that sequences related to GATCC are found 8-10 b preceding the TATA box of a number of genes: no function has been suggested for this conserved motif.

Approximately 80 b 5' to the transcribed region of the TK gene is the sequence GGCGAATTC. This sequence, which contains an EcoR1 recognition site, is related to the canonical sequence GGPyCAATCT that is found in the

-70 to -80b position in a number of RNA polymerase II - transcribed genes (24). Cleavage of HSV-1 DNA with EcoR1 greatly reduces the level of TK- transformation (27): moreover, deletions of the 3·4kb TK-containing BamH1 DNA fragment that remove the -80b sequence result in greatly reduced expression of TK activity in frog oocytes (28). Under these conditions, it appears that this conserved sequence may be part of a modulator of TK-gene expression. The -70b to -80b region of the IEmRNA-5 gene contains no such sequence. Indeed, this region forms part of a 42 bp sequence (residues 509-550) consisting entirely of G:C pairs. However, approximately 110 b 5' to the IEmRNA-5 start (residues 476-484) is the sequence GGCGAAGTG, 7 residues out of 9 of which correspond to the TK -80b sequence. It may be significant that this latter sequence forms part of a 10 bp perfect inverted repeat (residues 443-452 and 478-487) that is centered around a rather symmetric sequence of 18 A:T pairs (residues 457-474). These sequences lie within the Sma1 D subfragment of B4R1, within which the presence of two further small inverted repeats has already been noted. The HSV-1 TK gene contains a symmetric A:T paired region of 10 bp at around position -140b. However, no further regions of inverse symmetry are apparent in the vicinity.

It has been noted that sequences well upstream of the mRNA start modulate transcription of the sea urchin H2A histone gene (22). As yet, however, there is no evidence to implicate the unusual features in the 5' untranscribed region of the IEmRNA-5 gene with modulation of transcription. Approximately 800 bp separate the IEmRNA-5 start and that of IEmRNA-3, which is transcribed from the opposite DNA strand (5,6,29). This intervening region does not appear to be transcribed at any stage of the virus replicative cycle (unpublished results). As such, this region is a candidate to contain the DNA replication origin whose presence in TR_S/IR_S may be implicated by the sequences comprising the semi-autonomously replicating defective DNA (14,30).

Localization of the 3' terminus of IE mRNA-5: The 3' terminus of IEmRNA-5 was localized using an approach similar to that described above for the 5' terminus. From previous results (17) the IEmRNA-5 3' terminus was placed 570 b from the BamH1 site in the SalB5 fragment and it was inferred that the polyadenylation site mapped in a 550 bp Taq1 subfragment of SalB5. This Taq1 fragment was ^{32}P-labeled at the 3' termini by addition of α-^{32}P-dCTP using the Klenow fragment of E. coli DNA polymerase. Part of the sample was hybridized to cytoplasmic immediate-early RNA and the hybrids were treated with nuclease S1. The remainder of the sample was recut with MboII, the larger fragment was reisolated and subjected to the various sequencing reactions.

The products of nuclease digestion were then electrophoresed in parallel to the products of the sequencing reaction. A series of six predominant bands varying in size by increments of one nucleotide were obtained following nuclease S1 digestion (Figure 4). The smallest of this series corresponded in size to the T complementary to A at residue 2492 (Figure 2). This A is immediately preceded by a C that is 10 b from the sequence AATAAA (residues 2476-2481), a motif that has been recognized as forming an essential part of the polyadenylation signal (31). We expect that this sequence corresponds

Figure 4. Identification of the polyadenylation site of the IEmRNA-5 gene. The products of S1 digestion of hybrids formed between a 3' terminally labeled TaqI subfragment of SalB5 were electrophoresed in parallel to the sequence reaction products of the Taq I fragment.

to the polyadenylation signal,at least in part. The series of bands obtained is a probable consequence of the commonly noted incomplete digestion of single-stranded DNA by nuclease S1. From the map positions of the 3' and 5' termini of IEmRNA-5 it was determined that the transcribed region is approximately 1750 b. Considering an additional polyadenylate tail of approximately 150 b (32), it is apparent that this size determination of IEmRNA-5 is in reasonable agreement to the 2.0 kb estimate obtained by gel electrophoresis (4).

Cotranscript size and coding sequences of IE mRNA-5: The localization and sequence of the single IE mRNA-5 gene intron has been described previously (16). The gene intron is 149 bp in size and corresponds to residues 838-986 in Figure 2. From the co-ordinates of the intron and the presumed mRNA termini, the size of the IEmRNA-5 5' cotranscript should be 245 or 247 b and that of the 3' cotranscript 1505 b.

A search of the 5' cotranscript sequence did not reveal potential initiator codons, the first of which (ATG) is located in the 3' cotranscript 6b into U_S sequences (residues 1057-1059). This ATG is located 17 b from the sequence CTTCC (residues 1035-1039) complementary to a region of the 3' terminus of 18S rRNA suggested to be involved in binding mRNA (33). From this first ATG the reading frame is open for a further 87 codons. Translation of this sequence would result in synthesis of a polypeptide of molecular weight 9795, the predicted sequence of which is represented in Figure 2. Synthesis of an immediate-early polypeptide of estimated molecular weight 12000 was found to be specified by size fractionated RNA containing a mixture of IEmRNA-2, IEmRNA-4 and IEmRNA-5 (4). As the products of the first two mRNA's are known to be of 63000 and 68000 molecular weight (4,6), it is assumed that IEmRNA-5 specifies the 12000 molecular weight polypeptide. It should be noted that the size estimate of this small in vitro translated polypeptide is subject to error as its electrophoretic mobility corresponds to a region of the gel distorted by endogenous globin.

A search of the DNA sequence also revealed a second longer open reading frame in IEmRNA-5. This reading frame was open for 316 codons and extended from residues 1512 to 2456, terminating a few nucleotides preceding the polyadenylation signal. Assuming that the ATG nearest the mRNA 5' terminus is that recognized for protein synthesis initiation (34), use of this other open reading frame would require either that an additional splice is made to remove intervening termination signals or that a separate mRNA is transcribed with a different 5' terminus. Such mRNA's have not been detected.

We suggested previously that the sequences common to IEmRNA-5 and IEmRNA-4 mapping in TR_S/IR_S may comprise part of a coding sequence (17). The rationale behind this conclusion was that the unique sequences of IEmRNA-4, estimated to be less than 1450 b, were apparently too small to encode the 68000 molecular weight product of this mRNA (4,6). We show here that the reiterated sequences of IEmRNA-5 contain no initiation codon and indeed the first initiation codon of IEmRNA-4 maps in U_S, approximately 40b from the IR_S/U_S junction (Watson, R. J., unpublished results).

ACKNOWLEDGEMENTS

We thank Jake Maizel, Mike Dobson and Caroline Tolstohshev for computer analyses of this DNA sequence.

REFERENCES

1. Frenkel, N. and Roizman, B. (1972) Proc. Natl. Acad. Sci. U.S.A. 69: 2654-2658.
2. Swanstrom, R.I. and Wagner, E.K. (1974) Virology, 60: 522-533.
3. Clements, J.B., Watson, R.J. and Wilkie, N.M. (1977) Cell 12, 275-285.
4. Watson, R.J., Preston, C.M. and Clements, J.B. (1979) J. Virology 31: 42-52.
5. Clements, J.B., McLauchlan, J. and McGeoch, D.J. (1979) Nucleic Acids Research 7: 77-91.
6. Anderson, K.P., Costa, R.H., Holland, L.E. and Wagner, E.K. (1980) J. Virology 34: 9-27.
7. Watson, R.J. and Clements, J.B. (1980) Nature 280: 329-330.
8. Jones, P.C. and Roizman, B. (1979) J. Virology 31: 299-314.
9. Holland, L.E., Anderson, K.P., Shipman, C.,Jr. and Wagner, E.K. (1980) Virology 101: 10-24.
10. McKnight, S.L. (1980) Nucleic Acids Research 24: 5949-5964.
11. Wagner, M.J., Sharp, J.A. and Summers, W.C. (1981) Proc. Natl. Acad. Sci. U.S.A. 78: 1441-1445.
12. Leung, W.-C., Dimock, K., Smiley, J.R. and Bacchetti, S. (1980) J. Virology 36: 361-365.
13. Post, L.E., Mackem, S. and Roizman, B. (1981) Cell 555-565.
14. Denniston, K., Madden, M.J., Enquist, L.W. and Vande Woude, G.F. (1981) Gene, in press.
15. Enquist, L.W., Madden, M.J., Schiop-Stansly, P. and Vande Woude, G.F. (1979) Science 203: 541-544.
16. Watson, R.J., Umene, K. and Enquist, L.W. (1981) Nucleic Acids Research 9: 4189-4199.
17. Watson, R.J., Sullivan, M. and Vande Woude, G.F. (1981) J. Virology 37: 431-444.
18. Maxam, A. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol.65: 43-62, Academic Press, New York.
19. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18: 485-499.
20. Hentschel, C., Irminger, J.-C., Bucher, P. and Birnstiel, M.L. (1980) Nature 285: 147-151.
21. Green, M.R. and Roeder, R.G. (1980) Cell 22: 231-242.
22. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77: 7102-7106.

23. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77: 1432-1436.
24. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Research 8: 127-142.
25. Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980) Nucleic Acids Research 8: 957-977.
26. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P., (1980) Science 209: 1406-1414.
27. Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y-C. and Axel, R. (1977) Cell 11: 223-232.
28. McKnight, S.L. and Gavis, E.R. (1980) Nucleic Acids Research 24: 5931-5948.
29. Mackem, S. and Roizman, B. (1980) Proc. Natl. Acad. Sci. U.S.A. 77: 7122-7126.
30. Vlazny, D.A. & Frenkel, N. (1981) Proc. Natl. Acad. Sci. U.S.A. 78: 742-746.
31. Fitzgerald, M. and Shenk, T. (1981) Cell 24: 251-260.
32. Silverstein, S., Millette, R., Jones, P., and Roizman, B. (1976) J. Virology 18: 977-991.
33. Hagenbuchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) Cell 13: 551-563.
34. Kozak, M. (1978) Cell 15: 1109-1123.
35. McKnight, S.L., Gavis, E.R., Kingsbury, R. and Axel, R. (1981) Cell 25: 385-398.